DOI:
10.1021/acs.jnatprod.7b00377

Document Version
Peer reviewed version

Crispenes F and G, *cis*-Clerodane Furanoditerpenoids from *Tinospora crispa*, Inhibit STAT3 Dimerization

Md. Abdullah Al Noman,[†] Tasnova Hossain,[†] Monira Ahsan,[†] Shirin Jamshidi,[‡] Choudhury Mahmood Hasan*[†] and Khondaker Miraz Rahman*[‡]

[†]*Department of Pharmaceutical Chemistry, University of Dhaka, Dhaka-1000, Bangladesh*

[‡]*School of Cancer and Pharmaceutical Science, King's College London, 150 Stamford Street, London SE1 9NH, U K*

ABSTRACT

Two new *cis*-clerodane-type furanoditerpenes, crispenes F and G (**1** and **2**), together with seven known compounds, were isolated from the stems of *Tinospora crispa.* Crispenes F and G (**1** and **2**) inhibited STAT3 dimerization in a cell-free fluorescent polarization (FP) assay, and were found to have significant cytotoxicity against a STAT3-dependent MDA-MB 231 breast cancer cell line, while being inactive in a STAT3-null A4 cell line. These two compounds share structural similarities with a previously reported STAT3 inhibitor, crispene E, isolated from the same plant. Molecular docking studies suggested that the molecules inhibit STAT3 by interacting with its SH2 domain.

Oncogenic transcription factors are an increasingly important target for anticancer therapies, as their inhibition could allow for the "reprogramming" of tumor cells, leading to apoptosis or differentiation from the malignant phenotype.[1,2] Signal Transducer and Activator of Transcription Protein 3 (STAT3) has emerged as a promising molecular target for chemotherapy.[3] In vivo studies have demonstrated that STAT3 is constitutively active in a variety of malignancies, ranging from breast, prostate, and head and neck tumors to multiple myelomas and hematological cancers.[4] Cancer cells are often dependent upon the activation of STAT3, but non-cancerous cells are fairly tolerant of the loss of STAT3 function, likely reflecting redundancies in normal signal transduction.[5] This suggests a high therapeutic potential for STAT3 inhibitors. Furthermore, targeting STAT3 could potentially eliminate chemoresistance, as a number of signaling pathways converge on STATs and the inhibition of these proteins may forestall resistance.[6] Several growth factors or cytokines stimulate the STAT3 signaling pathway, which leads to receptor dimerization and activation. Phosphorylation of the tail of the receptor creates a docking site for the recruitment of an un-phosphorylated STAT3 (uSTAT3), which becomes phosphorylated at the Tyr705 position (near the C-terminus) by JAK kinases. The phosphorylation of the STAT3 protein leads to the formation of a homodimer through the reciprocal binding of the SH2 domain of one monomer to the pTyr-containing PYLKTK sequence of another. This dimeric pSTAT3 acts as a transcription factor by translocating to the nucleus, where it binds to its DNA consensus sequence, thus regulating the transcription of numerous genes critical for the survival and proliferation of cancer cells.[7,8] A number of different approaches have been taken to identify novel small molecules as STAT3:STAT3 dimerization inhibitors, including the development of peptides or peptidomimetics that bind to the STAT3 SH2 domain,[9] and small molecules, derived from STAT3 structural information and in silico design, which interact with the key

residues within the STAT3 SH2 domain.[10-12] A number of plant-derived molecules have also shown potent inhibition of the STAT3 signalling pathway and STAT3 dimerization.[13-16]

Historically, plants have represented an important pool for the identification of novel drug leads. There is increasing interest in searching for natural products with drug-like properties as potential preclinical candidates.[17] A fluorescent polarization assay has been used to study plant isolates with drug-like properties, including the previously reported crispene E, a *cis*-clerodane diterpene, as a STAT3 dimerization inhibitor.[16] As part of ongoing work to identify bio active plant metabolites from Bangladeshi medicinal plants,[16, 18-20] crispene F (**1**) and crispene G (**2**), two structurally related furanoid diterpenes with STAT3-inhibitory activity, were identified from *Tinospora crispa* Diels (Menispermaceae), a woody climber native to Malaysia, Indochina, the Indian subcontinent, and mainland China.[21] In traditional medicine, this species is used for the treatment of hypertension, diabetes mellitus, malaria, diarrhea, and as a vermifuge.[22, 23] All three of these diterpenes (**1**, **2** and crispene E) with a STAT3 dimerization inhibition property are drug-like, according to Lipinski's rule of five,[24] and the core scaffold is amenable for further medicinal-chemistry optimization. This may provide opportunities to enhance the STAT3 inhibitory activity and cytotoxicity of the scaffold in STAT3-dependent tumors.

RESULTS AND DISCUSSION

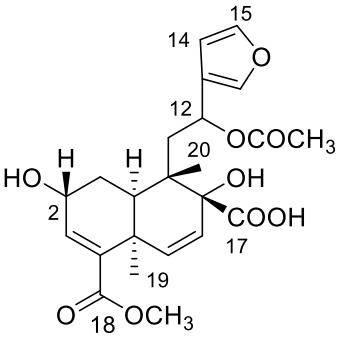
A methanol extract of the stem of *Tinospora crispa* led to the isolation of four clerodane-type furanoditerpene lactones (**1-4**), an alkaloid (**5**) and three terpenoids (**6-8**), and a sterol (**9**). Among the isolated diterpenes, compounds **1** and **2** are new, with a carboxylic acid group at position C-8 as found previously only in crispene C.[25] The known compounds were

identified as crispene D (**3**),[25] columbin (**4**),[26] *N*-*trans*-feruloyltyramine (**5**),[27] cycloeucalenone (**6**),[28] β-amyrin (**7**),[29] lupeol (**8**), and β-sitosterol (**9**), by comparison of their spectroscopic data with those reported in the literature and by co-TLC with authentic samples.

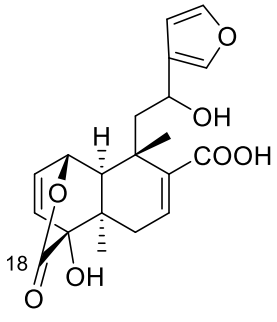
Table 1. NMR Spectroscopic Data (*J* in Hz) for Crispenes F (**1**)[a] and G (**2**)[b]

	crispene F (**1**)			crispene G (**2**)	
position	δ_C	δ_H (*J*in Hz)	HMBC	δ_C	δ_H (*J*in Hz)
1	28.6	2.01, ddd (14.6, 11.5, 6.7), 2.29, dd (14.6, 7.0)	2, 3, 5, 9, 10	74.9	5.19, br s
2	64.7	4.55, dd (11.5, 7.0)	3, 4	129.5	6.49, dd (7.2, 4.4)
3	139.7	6.71, s	1, 5, 18	136.8	6.40, d (7.2)
4	137.0			80.3	
5	37.7			36.9	
6	138.2	6.56, s	5, 8, 10	26.6	1.96, d (16) 2.48, dd (16, 8.0)
7	121.2	6.56 s	5, 8, 17	141.4	7.14, br d (8.0)
8	78.4			137.7	
9	39.5			42.3	
10	45.6	2.52, d (6.1)	1, 5, 9, 19, 20	44.7	1.72, br s
11	39.8	1.96, dd (13.5, 8.5) 2.25, dd (13.5, 8.5)	8, 9, 12, 13, 20	55.3	2.07, br d (5.6)
12	71.1	5.59, br t (8.5)	13, 14	69.9	4.93, br t (6.4)
13	126.0			123.3	
14	108.6	6.40, br s	15, 16	108.6	6.45, br s
15	144.3	7.43, br s	13, 16	143.9	7.43, br s
16	139.6	7.46, br s	13, 14, 15	140.1	7.50, br s
17	167.4			166.6	
18	166.9			175.3	
19	30.6	1.55, 3H s	5, 6, 10	25.5	1.13, 3H s
20	23.4	0.95, 3H s	8, 9, 10, 11	30.6	1.20, 3H s
-OCH_3	52.1	3.73, 3H s	18		
O$\underline{C}$OCH_3	168.8				
OCO$\underline{C}H_3$	21.3	2.11, 3H s	OC=O		

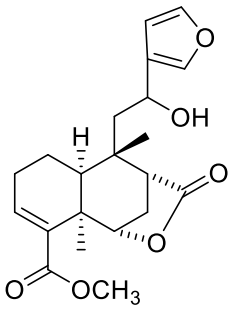
[a]In $CDCl_3$; ^1H NMR in 600 MHz; ^{13}C NMR in 150 MHz; [b] In $CDCl_3$; ^1H NMR in 400 MHz; ^{13}C NMR in 100 MHz.



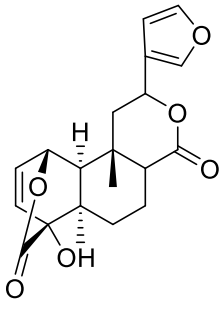
1



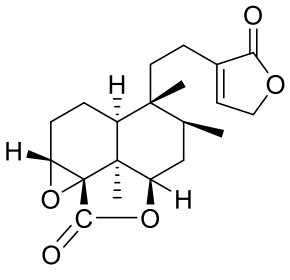
2



3



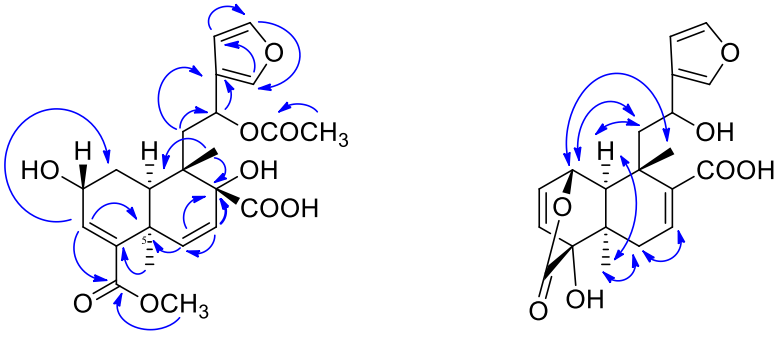
4



Crispene E

Compound **1**, isolated as colorless needles, gave the molecular formula $C_{23}H_{28}O_9$, as determined by HRESIMS measured in the negative-ion mode (*m/z* 447.1667 $[M-H]^-$). The ^1H NMR spectrum (Table 1) displayed resonances for three olefinic protons at δ_H 6.40, 7.43 and 7.46 (1H, br s each), assignable to protons of a furan ring as commonly observed in clerodane-type furanoid diterpenes isolated from *Tinospora* species.[30] Two characteristic angular methyl groups were observed at δ_H 1.55 (3H, s, Me-19) and δ_H 0.95 (3H, s, Me-20). A singlet at δ 6.71 (1H br s) was assigned to the C-3 proton of the trisubstituted double bond. This was confirmed by HMBC correlations from H-3 to C-1 and C-5 (Table 1, Figure 1). A broad olefinic proton signal at δ 6.56, which integrated for two protons (H-6 and H-7), showed direct correlations to C-6 and C-7 in the HSQC experiment and $^{2/3}J$ correlations to C-5 and C-8 in the HMBC spectrum. The ^1H and ^{13}C NMR spectra further revealed the presence of two oxymethine protons (δ_H 4.55, δ_C 64.7, H-2; δ_H 5.59, δ_C 71.1, H-12), an acetoxy methyl (δ_H 2.11, δ_C 21.3 and δ_C 168.8), and an ester methyl (δ_H 3.73, δ_C 52.1 and δ_C

166.9). In addition, the ^{13}C NMR spectrum revealed the presence of a carboxylic acid carbonyl group at δ_C 167.4, which was assigned as C-17. The NOESY spectrum showed correlations between H-2/H-20 and H-10 /H-19, suggesting OH-2 and H-10 and Me-19 as α and H-2 and Me-20 as β. From the spectroscopic data obtained, **1** was characterized as (2*R*,5*R*,9*S*,10*S*)-15,16 epoxy-2,8α-dihydroxy-12-acetoxycleroda-3,6,13(16),14-tetraen-18-methoxycarbonyl-17-oic acid, a new diterpene, which was given the trivial name crispene F.



1 **2**

Figure 1. Key HMBC correlations for compound **1** and NOESY correlations for compound **2**.

Compound **2** was also isolated as colorless needles, and its molecular formula was determined to be $C_{20}H_{22}O_7$ by HRESIMS, as measured in the negative-ion mode (*m/z* 373.1305 $[M-H]^-$). The ^1H and ^{13}C NMR spectra of compound **2** (Table 2) indicated the presence of a furan ring (δ_H 6.45, 7.43 and 7.50, 1H br s each; δ_C 108.6, 143.9 and 140.1) and two angular methyl groups (δ_H 1.13, 1.20 3H s, each; δ_C 25.5, 30.6), similar to compound **1**. Three olefinic protons at δ_H 6.49 (dd, J = 7.2, 4.4 Hz), 6.40 (d, J = 7.2 Hz), and 7.14 (br d, J = 8.0 Hz), and four unsaturated carbon resonances at δ_C 129.5, 136.8, 141.4, and 137.7 indicated a disubstituted and a trisubstituted olefinic bond assignable to H-2, H-3 and H-7, respectively. This was supported by the COSY spectrum, where H-2 coupled to H-3, and H-7 coupled to both the equatorial and axial protons of C-6. A broad triplet at δ 4.93 indicated an oxymethine proton and was assigned at C-12. The upfield shift of this proton suggested a free

hydroxy group at C-12 (δ_C 69.9). The ^{13}C NMR spectrum further revealed the presence of two oxygen-bearing carbons at δ_C 74.9 and 80.3, together with a ketone carbonyl group at δ_C 175.3, which indicated a lactone ring located between C-1 and an OH-bearing C-4, similar to that of columbin (**4**).[26] This was also supported by the close similarities of the ^{1}H and ^{13}C NMR shifts of ring A of these two compounds, in particular C-1 to C-6, C-10 and C-18. The HMBC spectrum for **4** was obtained, where the H-1 proton showed strong correlations to C-2, C-3, C-5, and C-18, confirming the position of a lactone ring between C-1 and C-4. The placement of the carboxylic acid group (δ_C 166.6) was fixed at C-8 on biogenetic grounds. The relative configuration was determined by a NOESY experiment, which showed correlations between H-10/H-19H-6α/H-14, H-1/H-20, H-11/H-20, and H-6/H-20 (Figure 1). On the basis of these data, the structure of compound **2** was assigned as 1*R*, 4*S*,5*R*,9*S*,10*S*,12*S*)-15,16-epoxy-4,12-dihydroxy cleroda-2,7,13(16),14-tetraen-18,1-olide, a new fuaranoditerpene named crispene G.

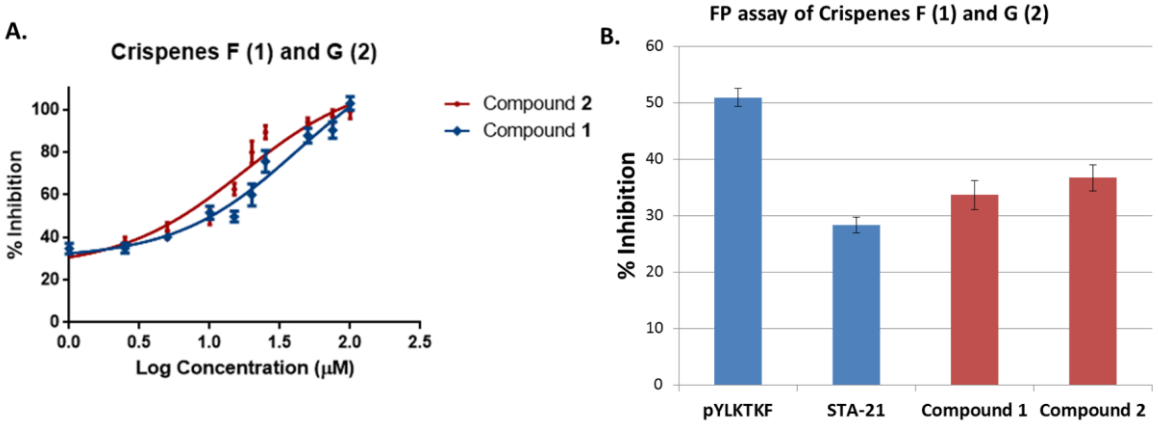


Figure 2. FP assay inhibition graphs for **1** and **2**. (a) PPI IC_{50} graph for **1** and **2**. (b) Inhibition relative to natural hexapeptide pYLKTKF and STA 21.

The unphosphorylated STAT3βtc protein was expressed in *E. coli (*BL21 Rosetta cells) and purified using an ion-exchange chromatographic protocol. Protein expression and purification protocols for U-STAT3βtc were adapted from a previously reported method.[31,32] The purified protein was used to develop a previously reported cell free fluorescent polarization (FP)-based primary protein-protein binding assay,[16] to assess the STAT3 dimerization inhibitory activity of the new compounds **1** and **2**. The FP assay was carried out using the fluorescein-labeled FAM-pYLPQTV peptide as the surrogate peptide and pYLKTKF peptide as a control inhibitor peptide to measure the inhibitory activity of **1** and **2**. Like crispene E, **1** and **2** disrupted STAT3 binding to the phosphorylated high-affinity peptide pYLPQTV-NH_2, with IC_{50} values of 42 μM and 17 μM, respectively. Compounds **1** and **2** showed 60% and 72% dimerization inhibition relative to pYLKTKF, respectively, and 119% to 130% inhibition relative to the SH2 domain interacting molecule STA-21[33] at 100 μM (Figure 2). The compounds were found to be less active compared to crispene E, which showed an IC_{50} of 10.3 μM and 210% inhibition relative to the SH2 domain interacting molecule STA-21.[16]

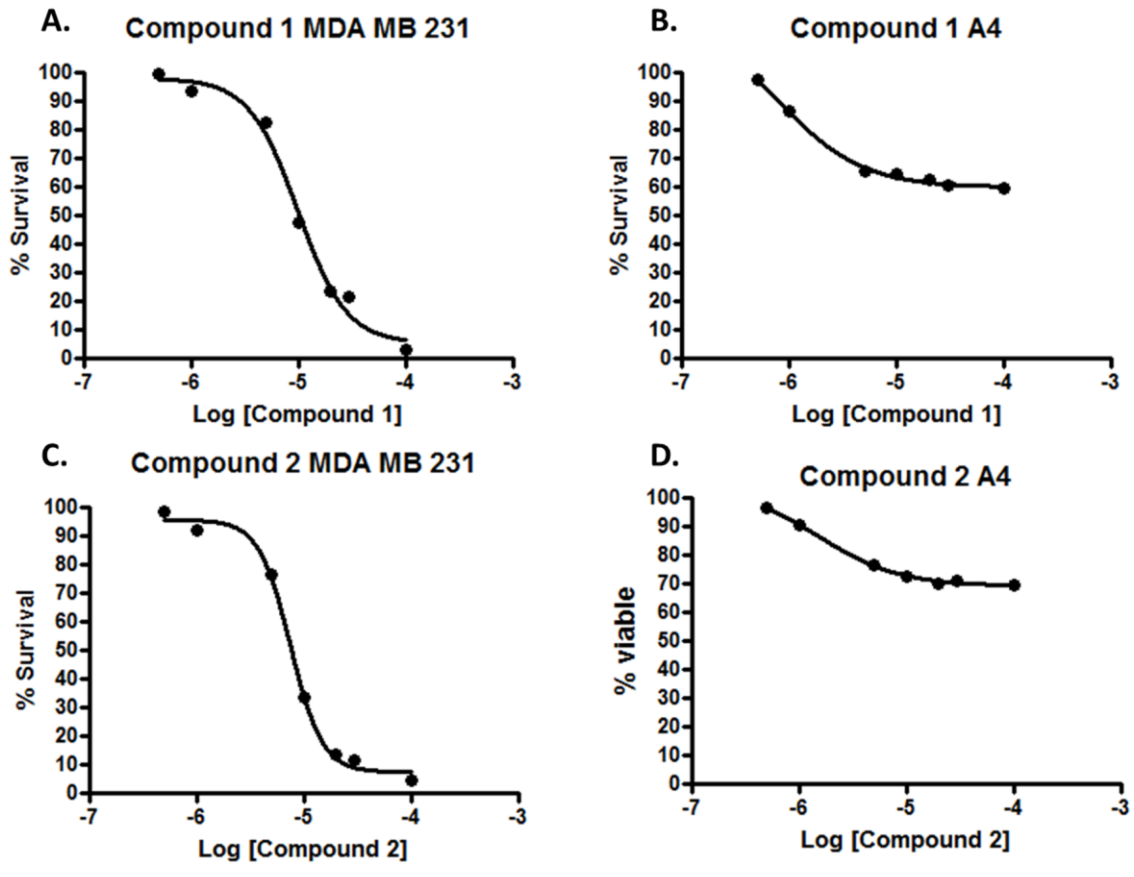


Figure 3: MTT cell-viability assay profile in MDA-MB-231 (STAT3-dependent) and A4 (STAT3-null) cells treated with **1** and **2** for 24 h.

To evaluate the effect of STAT3 dimerization inhibition on the viability of STAT3-dependent tumor cell lines and selectivity against non-tumor cell lines, **1** and **2** were tested using an MTT cell viability assay against both the MDA-MB-231 breast cancer (STAT3-dependent) and A4 (STAT3 null) colon cancer cell lines. Compound **1** showed an IC_{50} value of 10 µM against MDA-MB-231 cells, and >10 µM against A4 cells, while **2** showed an IC_{50} value of 7.8 µM against MDA MB 231 cells, and, like compound **1**, >10 µM against the A4 cell line, suggesting a STAT3-specific inhibition (Figure 3). However, both **1** and **2** showed some toxicity towards the STAT3-null A4 cell line at higher concentrations, and their IC_{50} values could not be obtained. The cytotoxicity results of the compounds were consistent with

the biophysical data, as compound **2** consistently showed greater activity compared to **1** in both the two cell-free STAT3 dimerization inhibition assay and the cellular assay, suggesting the potential role of the lactone ring in the interaction of **2** with the STAT3 protein. Interestingly, these two compounds were notably less active when compared to crispene E in both the FP assay and the MTT cytotoxicity assay, which showed an IC_{50} of 5.4 µM against the MDA-MB-231 cell line.[16] The compounds were also tested against a non-tumor lung fibroblast cell line, WI-38, but did not show any toxicity at the highest concentration (100 µM) tested (Supporting Information). The selective toxicity of the compounds against the tumor cell line is encouraging, as it provides a chemical scaffold that could be optimized with the help of a medicinal chemistry program.

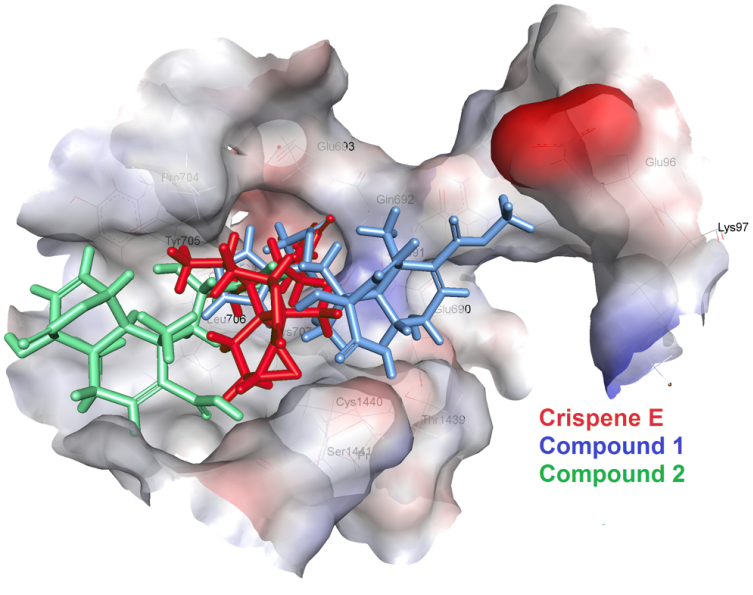


Figure 4. GOLD molecular docking poses of the compounds to the SMINA identified binding pocket of compounds **1** and **2** within the SH2 domain of STAT3.

A molecular modeling study was carried out to better understand the molecular mechanism of action of **1** and **2**, and the differences in activity compared to crispene E. Blind molecular docking with SMINA showed that the molecules bind to the adjacent

binding sites with the SH2 domain of the STAT3 protein. GOLD molecular docking of the molecules to the SMINA-identified binding pocket (Figure 4) showed compound **2** as the better molecule, with a GOLD score of 20.861 and an affinity value of –23.02 kcal/mol. The binding affinity values for crispene E and compound **1** were comparable, -21.74 and -21.50, respectively, but crispene E showed a notably superior GOLD score of 17.845 compared to 14.370 observed for **1**. The 2D interaction maps for **1** and **2** with their binding sites within the SH2 domain are shown in Figure 5 and in Table S1 (Supporting Information). Compound **1** formed hydrogen bonds with lysine 707 and serine 721 of the SH2 domain, and hydrophobic interactions were observed with proline 704 and leucine 706 for compound **1** and **2**. However, a hydrogen bond between compound **2** and Tyr 705, a critical residue for STAT3 dimerization, was observed. Interestingly, this observation was in line with the FP assay results observed for the compounds, as **2** showed greater STAT3 dimerization inhibitory activity. Interestingly, the binding mode of **1** and **2** to the pocket allows the molecule to directly inhibit the interaction of both Tyr705 and Leu706 of homodimer B. All of three ligands were in good interaction with the Hexad part of SH2 binding site residues. However, compound **2**, in addition to having good interactions with Hexad, interacted with the key Gln1355 (Gln633B) and Lys1348 (Lys626B) residues of the channel. This mechanism of action is in accord with previously published studies, describing molecules that prevent the interaction of Tyr705 in a similar manner.[33,34]

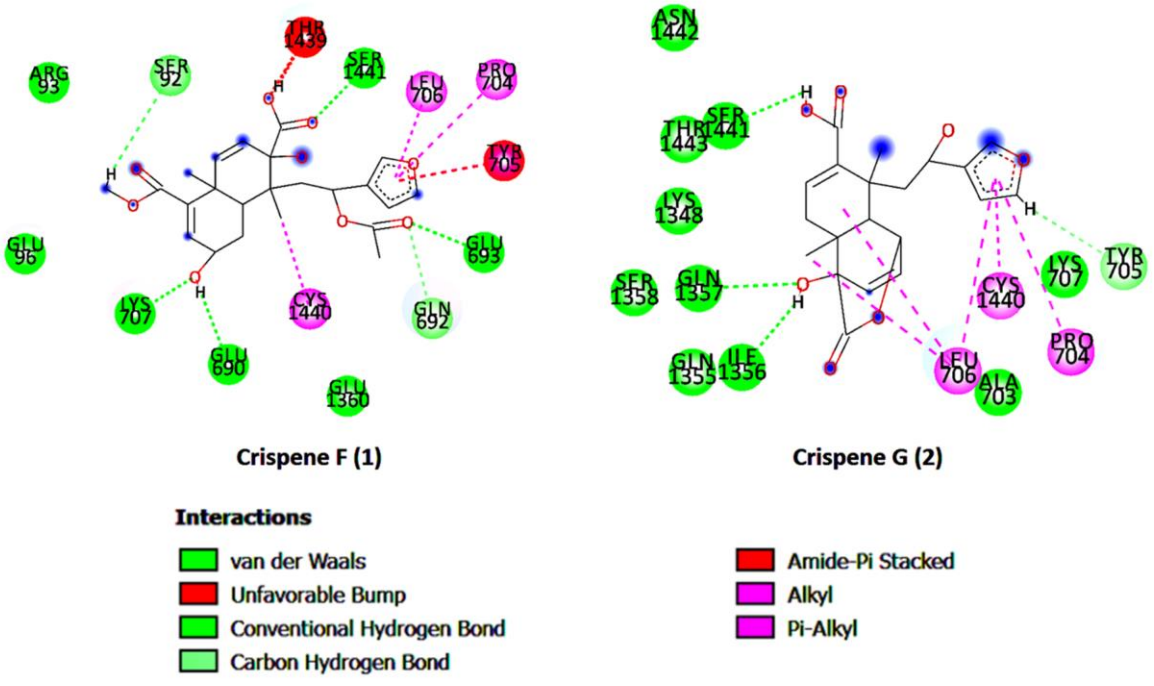


Figure 5. 2D interaction maps of **1** and **2** with the SH2 domain of STAT3 protein.

In conclusion, two new furanoid diterpenes (**1** and **2**) have been isolated from *T. crispa* and fully characterized structurally using spectroscopic techniques. These compounds showed notable STAT3 dimerization inhibitory activity in a cell-free primary protein-protein interaction assay, and showed selective toxicity towards a STAT3-dependent breast cancer cell line. Like previously reported structurally similar diterpenes from the same plant, these compounds appear to inhibit the STAT3 dimerization event by interacting with the Tyr705 of the SH2 domain. This confirms the STAT3-dimerization inhibitory activity of this diterpene scaffold.

EXPERIMENTAL SECTION

General Experimental Procedures. Optical rotations were recorded on a Bellingham Stanley digital polarimeter (London, U.K.). The optical rotation values were registered using chloroform as a solvent. NMR spectra were measured at 400 MHz for ^{1}H NMR spectra and 100 MHz for ^{13}C on a Bruker 400^{TM} ASCEND spectrometers in $CDCl_3$. High-resolution mass spectra (HRMS) were obtained on a Thermo Navigator mass spectrometer coupled to LC using electrospray ionization (ES) and time-of-flight (TOF) mass spectrometry. Vacuum-liquid column chromatography was performed using silica gel 60 (Merck, 0.015−0.040 mm). Silica gel (200-300 mesh) and Sephadex LH-20 (Sigma) were used for column chromatography. Thin-layer chromatography (TLC) was performed on E. Merck silica gel 60 F_{254} plates (0.25 mm) and the TLC plates were visualized under UV light at 254 and 366 nm, and using vanillin in sulfuric acid and Dragendorff's reagents. All solvents were purchased from commercial sources and distilled before use. LC-MS, liquid chromatography coupled with a mass spectrometer, was also used to determine the mass of the isolated compounds. The LC-MS analysis was performed on a Waters Alliance 2695 separation system using water (solvent A) and acetonitrile (solvent B) as mobile phases. Formic acid was used to ensure acidic conditions throughout the analysis and was used at a concentration of 0.1%. A 10 min and 5-min gradient was used to analyze submitted samples (ESI). The flow rate for both was 0.5 mL/min; 200 µL were split over a zero dead volume T piece, which passed into the mass spectrometer. The wavelength range of the UV detector of the HPLC system was 220-500 nm. The LC system was equipped with a photodiode array (535 scans) as the detector and a monolithic C_{18} 50 × 4.60 mm column.

Plant Material. The stems of *Tinospora crispa* were collected in December 2014 from the Tangail district of Bangladesh. The plant was identified by Mr. Bushra Khan, of the Bangladesh National Herbarium (BNH; Dhaka, Bangladesh), where a voucher specimen was deposited (DACB accession number 40929). The stems were cleaned, cut into small pieces, and subsequently subjected to shade drying for a week. The plant material was then oven dried for 24 h at a low temperature (not more than 40 °C) for better grinding. The dried stem was then crushed into a coarse powder by a high capacity grinding machine with proper care.

Extraction and Isolation. Powdered plant material (850 g) was extracted three times with distilled methanol. The combined extracts were concentrated to dryness (51.5 g) under reduced pressure using a rotary evaporator at 40 °C temperature. The extract was separated into 46 fractions by vacuum-liquid chromatography (VLC) using mixtures of solvents (petroleum ether, CH_2Cl_2, EtOAc, and MeOH) of increasing polarity. VLC fractions 11, 15 and 27 (F_{11}, F_{15} and F_{27}) were further fractionated on a Sephadex LH-20 column with petroleum ether-chloroform (1:4, 1:8) and 100% chloroform sequentially. Subfraction 15 of F_{11} was subjected to preparative TLC over silica gel using EtOAc-toluene, 5:95, to obtain **6** (3.1 mg; R_f 0.71 in EtOAc-toluene, 5:95). Sub-fractions 10-12 of F_{15} were combined and subjected to preparative TLC using EtOAc-toluene, 8:92, to obtain **7** and **8** (as a mixture**)** and **9** (3.8 mg and 10 mg; R_f 0.66 and 0.59 in EtOAc-toluene, 5:95). Subfractions 24-28 of F_{27} were combined and crystals of **4** (24 mg; R_f 0.64 in EtOAc-toluene, 40:60) were obtained. The supernatant was allowed to recrystallize to obtain **1** (10.5 mg; R_f 0.37 in EtOAc-toluene, 40:60). After separation of the crystals of **1**, the mother liquor was subjected to preparative TLC over silica gel using EtOAc-toluene, 40:60, to obtain **2** and **3** (2.4 mg and 3.6 mg; R_f 0.53 and 0.66, respectively, in EtOAc-toluene, 40:60).

Crispene F ***(1)***. Colorless needles; $[\alpha]^{25}_{D}$ -380 (*c* 0.5, $CHCl_3$); ^{1}H NMR and ^{13}C NMR data, Table 1; HRESIMS *m/z* 447.1667 $[M - H]^-$ (calcd for $C_{23}H_{27}O_9$ – H, 447.1652).

Crispene G ***(2)***. Colorless needles; $[\alpha]^{25}_{D}$ -306 (*c* 0.5, $CHCl_3$); ^{1}H NMR and ^{13}C NMR data, Table 2; HRESIMS *m/z* 373.1305 $[M - H]^-$ (calcd for $C_{20}H_{21}O_7$ – H, 373.1285).

Molecular Modeling. The coordinates of PDB ID 1BG1 were used for generating the structure of the STAT3 homodimer protein. 1BG1 contains one chain of the dimer and there are some missing residues (1-135, 185−193, 689−701 and 717−770). The missing residues up to 722 and the second chain were prepared by a homology modeling approach using the Accelrys Discovery Studio 2.5. Finally, the obtained structure was minimized, equilibrated and simulated by the use of AMBER 12 to generate a more relaxed STAT3 structure that was used for molecular docking. Molecular docking was performed to generate several distinct binding orientations and a binding affinity for each binding mode. Subsequently, the lowest binding free energy was considered as the most favorable binding mode for the system. AutoDock SMINA,[36] which uses the AutoDock Vina scoring function by default, was used for the blind molecular docking of the ligands to the STAT3 structure (PDB ID: 1BG1) for finding the best binding site, by exploring all probable binding cavities of the proteins. SMINA was performed with default settings, which samples nine ligand conformations using the Vina docking routine of stochastic sampling. Then, GOLD molecular docking was applied for the docking of crispene E, and compounds **1** and **2** to the SMINA-located best binding site of the STAT3, SH2 binding domain, for performing flexible molecular docking, as described elsewhere,[37, 38] Based on the fitness function scores and ligand binding positions, the best-docked poses for the ligands were selected. The GOLD molecular docking procedure was performed by applying the GOLD protocol in the Accelrys Discovery Studio software.[39] The Genetic algorithm (GA) was used in GOLD ligand docking software to examine

thoroughly the ligand conformational flexibility along with the partial flexibility of the protein.[37] The maximum number of runs was set to 20 for the ligand and the default parameters selected were 100 population size, 5 for the number of islands, 100,000 for the number of operations, and 2 for the niche size. Default cut-off values of 2.5 Å (dH-X) for hydrogen bonds and 4.0 Å for the van-der-Waals distances were applied. When the top solutions attained the RMSD values within 1.5 Å, the GA docking was terminated.

Growth of *E. coli* Cells Containing the STAT3βtc Plasmid. An aliquot of the master cell bank (BL21 Rosetta cells transformed with the STAT3βtc plasmid; kindly provided by Müller and co-workers, European Molecular Biology Laboratory, France), was grown overnight at 37 °C in the presence of ampicillin and chloramphenicol. One of the seed cultures was used to inoculate a 10 L Electrolab fermenter in the presence of 100 µg ampicillin and 20 µg chloramphenicol. The seed was gently stirred in the fermenter at 250 rpm and 37 °C, with a suitable air flow (3 L/min). The cells were induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG), when an OD of 0.6 was reached. The temperature was then reduced to 21 °C, and the bacteria grown overnight before being harvested by centrifugation at 3600 × *g* for 25 min at 4 °C.

Extraction of the Unphosphorylated STAT3βtc Protein. To each gram of pellet, 10 mL of extraction buffer (20 mM Hepes-KOH, pH 7.6, 0.1 M KCl, 10% glycerol, 1 mM EDTA, 10 mM $MnCl_2$, 20 mM DTT, 0.5 mM PMSF, protease inhibitor cocktail tablet) was added to re-suspend the pellet. Each sample was placed into a chilled beaker of ice and sonicated for 5 min at 15 mA, repeating for 15 sec, or followed by 15 sec off, using a medium-sized sonicator probe (TS701 H, Camlab Trans-sonic, Cambridge, U.K.). The solution was then centrifuged (JA25.50 rotor, Beckman Coulter Centrifuge, Brea, CA, USA,

4 °C, 1 h, 27,000 × *g*), and the supernatant collected and chilled prior to protein precipitation using ammonium sulfate. The mixture was then centrifuged (JA25.50 rotor, Beckman Coulter centrifuge, 4 °C, 1 h, 27,000 × *g*), and the pellet stored at 4 °C until required.

Purification of the Unphosphorylated STAT3βtc Protein. The crude STAT3 protein pellet was dissolved in 5 mL salt-free buffer (100 mM Tris pH 8.5, 1 mM EDTA, 2 mM DTT), and the solution was filtered through a 0.22 μm filter. The filtrate was then diluted to 20 mL, using double distilled H_2O. Approximately 5 mL of the dissolved protein were passed through an ion-exchange (IE) column (GE Healthcare, Little Chalfont, Buckinghamshire, UK) and stored on ice during collection. The IE column was washed with 10 mL salt-free buffer, before eluting the protein from the column with successive volumes (20 mL) of each of the ion-exchange buffers, increasing the concentration of NaCl each time. The purified protein was dialyzed overnight in either 10 mM ammonium acetate (to collect ESIMS data) or 100 mM ammonium bicarbonate (for digestion purposes).

FP Assay Protocol. The FP assay was carried out using the purified uSTAT3 protein and FAM-labeled pYLPQTV in PBS buffer (pH 7.4) in a 96-well plate format using a Corning black plate. The assay volume was 200 μL in each well. The Corning black plate was thoroughly washed with distilled water and allowed to dry prior to the assay. A protein-surrogate peptide complex was formed in PBS buffer in each well by adding FAM-pYLPQTV to uSTAT3 protein, with a final concentration of 10 nM for the peptide and 350 nM for the protein. The MP value of the surrogate dimer complex was measured using a fluorescent plate reader (Envision, Perkin Elmer, Waltham, MA, USA). This provided the base MP value. At this point, the inhibitor solution was added to the well and the assay plate was left on the shaker for 5 min. The MP value for each well was again measured by the fluorescent plate reader. This represented a shift in fluorescent value due to the displacement

of the fluorescently-labeled surrogate peptide by the inhibitors, and the average inhibition for each inhibitor was calculated as follows:

Base MP value (X) for each well = MP value of Protein and Probe (350 nM STAT3 + 10 nM FAM-pYLPQTV) - MP value of free FAM-pYLPQTV

Inhibitor MP value (Y)for each well = MP value of Protein +Probe + Inhibitor (350 nM uSTAT3 + 10 nM FAM-pYLPQTV + 2 µL Inhibitor) - MP value of free FAM-pYLPQTV with 2 µL DMSO

% inhibition by any ligand/peptide = (X-Y)/X*100

For comparison purposes, the % inhibition produced by 100 µM LKTKFI was considered as 100%, and the inhibition produced by different ligands relative to 100 µM LKTKFI was measured.

% Inhibition produced by 100 µM LKTKFI = A

% Inhibition produced by 100 µM Ligand = B

Relative inhibition (%) = B/A*100

Cell Culture. The MDA-MB-231 (triple-negative breast cancer), A4 (colon cancer) and WI38 (lung fibroblast) cell lines were obtained from the American Type Culture Collection. The MDA MB 231 cell line was maintained in Dulbecco's modified Eagle's medium (DMEM; Invitrogen), supplemented with fetal bovine serum (10% v/v; Invitrogen), l-glutamine (2 mM; Invitrogen), non-essential amino acids (1x; Invitrogen) and penicillin-streptomycin (1% v/v, Invitrogen). The A4 cell line was maintained in RPMI 1640 medium supplemented with fetal bovine serum (10% v/v; Invitrogen) and penicillin-streptomycin (1% v/v, Invitrogen). The non-tumor cell line WI-38 was maintained in Dulbecco's modified

Eagle's medium (DMEM; Invitrogen), supplemented with fetal bovine serum (10% v/v; Invitrogen). During seeding, cells were counted using a Neubauer hemocytometer (Assistant, Hanover, Germany) by microscopy (Nikon, Melville, NY, USA) on a non-adherent suspension of cells that were washed in PBS, trypsinized, centrifuged at 8 °C at 8000 rpm for 5 min, and re-suspended in fresh medium.

MTT Assay. The cells were grown in normal cell culture conditions at 37 ºC under a 5% CO_2 humidified atmosphere using an appropriate medium. The cell count was adjusted to 10^5 cells/mL and 2,500 cells (MDA-MB-231) or 5,000 cells (A4 and WI-38) were added per well. The cells were incubated for 24 h and 1 μL of the appropriate inhibitor concentrations was added to the wells in triplicate. After 96 h of continuous exposure to each compound, the cytotoxicity was determined using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (Lancaster Synthesis Ltd, Morecambe, Lancashire, UK) colorimetric assay.[35] Absorbance was quantified by spectrophotometry at λ = 570 nm (Envision Plate Reader, Perkin Elmer, Waltham, MA, USA). IC_{50} values were calculated by a dose-response analysis using the Prism GraphPad Prism® software.

ASSOCIATED CONTENT

The Supporting Information is available free of charge on the ACS Publications website at DOI: XXX.XXX HPLC methods, molecular modeling images and ^{1}H and ^{13}C NMR spectra of isolated compounds (PDF).

AUTHOR INFORMATION

Corresponding Authors:

*Tel: +442078481891. Fax: +442078484295.. E-mail: k.miraz.rahman@kcl.ac.uk (K. M. Rahman).

*Tel: +8801819-253698. Fax: +88 02 9667222 . E-mail: cmhasan@gmail.com (C. M. Hasan).

ORCID

Khondaker Miraz Rahman: 0000-0001-8566-8648

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are thankful to Mr. Bushra Khan, of the Bangladesh National Herbarium (BNH; Dhaka, Bangladesh) for identifying the plant material, and Dr. John Anderson of the UCL Institute of Child Health for providing the A4 cell line. Dr. Fakir Shahidullah Tareq is acknowledged for running the 600 MHz NMR spectrum of compound **1** at the University of Bremen, Germany.

REFERENCES

(1) Luo, J.; Solimini, N. L.; Elledge, S. J. *Cell* **2009,** *136*, 823-837.

(2) Darnell, J. E. *Nat. Rev. Cancer* **2002,** *2*, 740-749.

(3) Turkson, J. *Expert Opin. Ther. Targ.* **2004,** *8*, 409-422.

(4) Turkson, J.; Jove, R. *Oncogene* **2000,** *19*, 6613-6626.

(5) Yue, P.; Turkson, J. *Expert Opin. Inv. Drugs* **2009,** *18*, 45-56.

(6) Haura, E. B.; Turkson, J.; Jove, R. *Nat. Clin Pract. Oncol.* **2005,** *2*, 315-324.

(7) Leonard, W. J.; O'Shea, J. J. *Annu. Rev. Immunol.* **1998,** *16*, 293-322.

(8) Levy, D. E.; Darnell, J. E. *Nat. Rev. Mol. Cell Biol.***2002,** *3*, 651-662.

(9) Turkson, J.; Kim, J. S.; Zhang, S. M.; Yuan, J.; Huang, M.; Glenn, M.; Haura, E.; Sebti, S.; Hamilton, A. D.; Jove, R. *Mol. Cancer Ther.* **2004,** *3*, 261-269.

(10) Bromberg, J.; Darnell, J. E. *Oncogene* **2000,** *19*, 2468-2473.

(11) Siddiquee, K.; Zhang, S.; Guida, W. C.; Blaskovich, M. A.; Greedy, B.; Lawrence, H. R.; Yip, M. L. R.; Jove, R.; McLaughlin, M. M.; Lawrence, N. J.; Sebti, S. M.; Turkson, J. *Proc. Natl. Acad. Sci. USA.* **2007,** *104*, 7391-7396.

(12) Berg, T. *Chembiochem* **2008,** *9*, 2039-2044.

(13) Nam, S.; Xie, J.; Perkins, A.; Ma, Y.; Yang, F.; Wu, J.; Wang, Y.; Xu, R.-Z.; Huang, W.; Horne, D. A. *Mol. Oncol.* **2012,** *6*, 484-493.

(14) Maloney, K. N.; Hao, W.; Xu, J.; Gibbons, J.; Hucul, J.; Roll, D.; Brady, S. F.; Schroeder, F. C.; Clardy, J. *Org. Lett.* **2006,** 8, 4067-4070.

(15) Siveen, K. S.; Sikka, S.; Surana, R.; Dai, X.; Zhang, J.; Kumar, A. P.; Tan, B. K.; Sethi, G.; Bishayee, A. *Biochim. Biophys. Acta, Rev. Cancer* **2014,** *1845*, 136-154.

(16) Mantaj, J.; Rahman, S. A.; Bokshi, B.; Hasan, C. M.; Jackson, P. J.; Parsons, R. B.; Rahman, K. M. *Org. Biomol. Chem.* **2015,** *13*, 3882-3886.

(17) Butler, M. S. *Nat. Prod. Rep.* **2008,** *25*, 475-516.

(18) Haque, M. R.; Rahman, K. M.; Iskander, M. N.; Hasan, C. M.; Rashid, M. A. *Phytochemistry* **2006,** *67*, 2663-2665.

(19) Sohrab, M. H.; Chowdhury, R.; Rahman, K. M.; Hasan, C. M.; Rashid, M. A. *Fitoterapia* **2004,** *75*, 510-513.

(20) Mottakin, A. K. M.; Chowdhury, R.; Haider, M. S.; Rahman, K. M.; Hasan, C. M.; Rashid, M. A. *Fitoterapia* **2004,** *75*, 355-359.

(21) Kadir, F. A.; Othman, F.; Abdulla, M. A.; Hussan, F.; Hassandarvish, P. *Indian J. Pharmacol.* **2011,** *43*, 64-68.

(22) Klangjareonchai, T.; Roongpisuthipong, C. *J. Biomed. Biotechnol.* **2012,** *2012*, 808762.

(23) Zaridah, M.; Idid, S.; Wan Omar, A.; Khozirah, S. *J. Ethnopharmacol.* **2001,** *78*, 79-84.

(24) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Deliver. Rev.* **2001,** *46*, 3-26.

(25) Hossen, F.; Ahasan, R.; Haque, M. R.; Begum, B.; Hasan, C. M. *Pharmacogn. Mag.* **2016,** *12*, S37.

(26) Waterman, P. G.; Hasan, C. M.; Jabbar, A *Planta Med.* **1985,** *51*, 181-182.

(27) Park, J. B. *J. Agri .Food Chem.* **2009,** *57*, 8868-8872.

(28) Kongkathip, N.; Dhumma-Upakorn, P.; Kongkathip, B.; Chawananoraset, K.; Sangchomkaeo, P.; Hatthakitpanichakul, S. *J. Ethnopharmacol.* **2002,** *83*, 95-99.

(29) Gealt, M. A. *Microbiology* **1983,** *129*, 543-546.

(30) Ahmad, S. *Phytochemistry* **1988,** *27*, 1882-1884.

(31) Heidelberger, S.; Zinzalla, G.; Antonow, D.; Essex, S.; Piku Basu, B.; Palmer, J.; Husby, J.; Jackson, P. J. M.; Rahman, K. M.; Wilderspin, A. F.; Zloh, M.; Thurston, D. E. *Bioorg. Med. Chem. Lett.* **2013,** *23*, 4719-4722.

(32) Nkansah, E.; Shah, R.; Collie, G. W.; Parkinson, G. N.; Palmer, J.; Rahman, K. M.; Bui, T. T.; Drake, A. F.; Husby, J.; Neidle, S.; Zinzalla, G.; Thurston, D. E.; Wilderspin, A. F. *FEBS Lett.* **2013,** *587*, 833-839.

(33) Song, H.; Wang, R.; Wang, S.; Lin, J. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 4700-4705.

(34) Page, B. D. G.; Fletcher, S.; Yue, P.; Li, Z.; Zhang, X.; Sharmeen, S.; Datti, A.; Wrana, J. L.; Trudel, S.; Schimmer, A. D.; Turkson, J.; Gunning, P. T. *Bioorg. Med. Chem. Lett.* **2011,** *21*, 5605-5609.

(35) Mosmann, T. *J. Immunol. Methods* **1983,** *65*, 55-63.

(36) Koes, D. .R.; Baumgartner, M. P.; Camacho, C. J. *J. Chem. Inf. Model*, **2013**, *53*, 1893-1904.

(37) Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995,** *245*, 43-53.

(38) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997,** *267*, 727-48.

(39). Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. *Proteins* **2002,** *49*, 457-471.